

DOCUMENT RESUME

ED 156 645

95

SP 012 783

AUTHOR Peck, Robert F.; And Others
 TITLE The Consistency of Individual Teaching Behavior.
 INSTITUTION Texas Univ., Austin. Research and Development Center
 for Teacher Education.
 SPONS AGENCY National Inst. of Education (DHEW), Washington,
 D.C.
 PUB DATE Mar 78
 CONTRACT NE-C-00-3-0066
 GRANT NIE-G-78-0116
 NOTE 13p.; Paper presented at the Annual Meeting, American
 Educational Research Association (Toronto, Canada,
 March 27-31 1978)
 EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
 DESCRIPTORS Academic Achievement; Effective Teaching; Elementary
 Education; Observation; *Personality Studies;
 *Student Evaluation of Teacher Performance; *Teacher
 Behavior; Teacher Characteristics; *Teacher
 Evaluation; Teaching Styles

ABSTRACT

The consistency of teaching behavior of 28 sixth-grade teachers over a period of two years was studied and evaluated. Trained observers rated teachers with Ryans' Classroom Observation Record (20 scales that reproduced Ryans' factors of teacher behaviors). Many of the same teachers were observed again in the following year by a team of half-old, half-new observers. The observers also rated student behavior on seven scales. An independent appraisal of the teachers by their students was carried out in both years. Different sets of adults and student judges, in two successive years, described these teachers with a substantial degree of consistency, even though the teachers taught different subjects in the two years. Findings indicate that there is a considerable amount of individual stability in important aspects of teaching behavior, at least across a two year span. (JD)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED156645

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

THE CONSISTENCY OF INDIVIDUAL TEACHING BEHAVIOR

Robert F. Peck, Nils Greger Olsson, Janet L. Green

Research and Development Center for Teacher Education

The University of Texas at Austin

Paper presented at the annual meeting of the American Educational Research Association held in Toronto, Canada, March 27-31, 1978, Program Session 18.26.

This study was supported in part by National Institute of Education Contract NE-C-00-3-0066, Grant Number OB-NIE-G-78-0116, Research and Development Center for Teacher Education. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that office should be inferred.

SP012 783

The Consistency of Individual Teaching Behavior

R. PECK, G. OLSSON, J. GREEN

Research and Development Center for Teacher Education

If a considerable part of teaching behavior is an expression of the teacher's personality and values, and if individual personality is persistent across time, it should be possible to demonstrate a great degree of individual consistency in teaching behavior if the right kinds of measures are taken, at the right levels of generality. This would seem to argue for fairly high inference measures which allow for different "phenotypic" expressions of stable "genotypic" traits. Different classroom situations may call for different "phenotypic" actions, such as low inference coding systems address.

The evidence of earlier studies has been equivocal; but a substantial and increasing amount of positive evidence indicates that some important dimensions of individual teaching behavior do tend to be consistent across time and contexts. Shavelson and Dempsey-Atwood (1976), in their review of existing stability data, concluded that not much consistency of teacher behavior has been shown, although Brophy and Evertson (1977) pointed out many methodological limitations in the studies they cited. A longitudinal study of over 1,000 elementary teachers, however, reports "a marked to very high correlation (0.42 for college advisor-pairs; 0.80 for first year by third year evaluator) between the six evaluators of the teacher from student teaching to the third year in the profession." (Myers, 1967) The evaluative scales have to be examined in more detail, of course, to see what kinds of teaching behavior were thus evaluated.

Simon (1967) and Pfeiffer (1966) also found considerable consistency in teachers' behavior from year to year, or across classes. Some other studies, using low inference measures, found more variation, often attributable to the

constraints of different subject-matter and different classroom structures (Marshall, et. al, 1976). Brophy and Evertson (1977), however, found consistency coefficients of 0.49 to 0.74, and higher, on numerous low-inference variables when junior high school teachers were compared across two different classes. High-inference ratings on the Emmer-Peck scales, based on the same observations, ranged from 0.73 to 0.86. An earlier study (Brophy, et. al, 1975) found high teacher stability, for most of these scales, across two school years.

In designing the current Teaching-Learning Interaction Study, provision was made to test the proposition of teacher consistency in several cross-checking ways. First, teaching behavior was rated by trained observers with Ryans' Classroom Observation Record (Ryans, 1960): twenty scales that reproduced Ryans' factors of Systematic, Organized (COR SO), Kindly, Understanding (COR KU), and Stimulating, Imaginative (COR SI) behaviors, in social studies classes, chiefly. Many of the same teachers were observed again in the following year by a team of half-old, half-new observers. However, both classes and subject area (reading) differed from the first year. In each year, the observers also rated student behavior on seven scales, yielding a factor score on "Positive Work Orientation" (COR Pupil) for each class. Finally, an independent appraisal of the teachers by their students was carried out in both years, using the Student Evaluation of Teaching (SET) instrument. Again, these ratings were based on different classes of students, in different subject areas, from year to year.

Sample

This sample represents a subset of the entire Austin sample used in the TLI study. The Austin sample was based on 53 classes of sixth grade students from four schools in the first year of the study, and 43 sixth grade classes

from three schools, in the second year. This analysis uses the data from 28 teachers who were common to both years of the study, and for whom all measures relevant to this study were available. Students were randomly assigned to classrooms, after assuring that classes were racially mixed in the same proportions as the Austin community as a whole (69.5% Anglos; 14.9% Blacks; and 15.6% Chicanos).

Instruments

Classroom Observation Record (COR)

The Classroom Observation Record (COR) uses trained observers to rate the behavior of teachers and pupils. The format consists of a series of bipolar adjectives, separated by a 7-point scale. The original instrument was developed by D.F. Ryans and published in Characteristics of Teachers (1960). Slightly modified versions, with a few items added to the scales, were employed in this study.

In both years, observers rated teacher and pupil behavior during 40-minute class periods, across an average of 2 to 5 occasions. Factor analyses of the mean scores (across both raters and occasions), for the items upon which teachers were rated, indicated a 3-factor structure which corresponded almost exactly with that obtained by Ryans. These factors were scored for each teacher and were given Ryans' names: "Kindly, Understanding" (COR KU), based on 9 items; "Systematic, Organized" (COR SO), based on 6 items; and "Stimulating, Inventive" (COR SI), based on 5 items. In addition, a pupil behavior (COR Pupil) score was derived for each teacher, based on observers' ratings of pupils, on 7 additional items. The rationale for inclusion of this scale is Ryans' (1960) finding that, based on sets of ratings made by the same observers, desirable pupil behavior correlated highly (0.78 to 0.83 in grades 1-6) with the teacher behavior factors.

Internal consistency reliability (Cronbach's alpha) for the four scales was 0.94, 0.96, 0.97, and 0.95 for Year 1 (N=53) and 0.93, 0.93, 0.98, and 0.88 for Year 2 (N=43) for COR Pupil, KU, SO, and SI respectively. Estimated rater reliability for the three teacher scales, computed from the independent ratings of 20 teachers by 4 observers on 2 occasions, was 0.97 (KU), 0.98 (SO), and

0.98 (SI). The Year 2 rater reliability, based on a team of half-old, half-new observers, was similar: 0.96 (KU), 0.99 (SO), and 0.97 (SI). Rater reliability for the COR Pupil factor in Year 2 was 0.96.

Student Evaluation of Teaching (SET)

The Student Evaluation of Teaching (SET) instrument provides an appraisal of teachers by their students. The original version of this instrument was developed during the 1960's by D.J. Veldman and R.F. Peck at the University of Texas at Austin Research and Development Center for Teacher Education (Veldman, 1970). Slightly modified versions were employed in this study. In all cases, students are asked to respond to a series of descriptive phrases about their teacher (e.g., "This teacher is friendly toward students" or "This teacher knows a lot about the subject") by rating them on a Likert scale.

The SET was administered to students at the end of each of the two years of the study. Analysis of the data indicated that for each year, the items employed measured a single underlying dimension. A mean score across all items, for all students in a class, was therefore used as the SET measure of "Teacher Effectiveness", in each year.

Internal consistency reliability (Cronbach's alpha) for this instrument was 0.80 (N=1297) for the first year and 0.81 (N=1128) for the second. Data for test-retest reliability was collected during the second year of the study and is 0.78 for a time lapse of two weeks (N=61). Evidence of the validity of the original instrument is given by Veldman (1970) in a study of 50 teachers, each having two classes of students. Correlations across classes for the five scales constructed from the SET (2 items per scale) ranged from 0.72 to 0.92.

Comprehensive Tests of Basic Skills (CTBS)

The Comprehensive Tests of Basic Skills (CTBS) was published by CTB/McGraw-Hill in 1973. It represents a series of standardized tests which measure

students' academic achievement in the areas of language arts, social studies, math, and science. For each content area there are several forms of these tests available, at several levels of difficulty. Published scoring procedures exist for comparing scores on different forms, at different levels of difficulty, and across content areas.

In the present study, the CTBS was administered at the beginning and end of the first year (pretest and posttest). Depending on the subject taught by their homeroom teacher, students were tested either in Language Arts, composed of language mechanics and language expression (6 classes), or Social Studies (22 classes). For the pretest the CTBS, Form S, Level 2 was utilized, and for the posttest, students received CTBS, Form S, Level 3.

Both internal consistency reliability (KR-20) and interlevel correlation coefficients are reported in the CTBS Technical Bulletin No. 1 (1974, p. 41-42) for each of the tests for a particular sample of students. For late fifth and early sixth graders (pretest level), the Form S, Level 2 Language Arts' internal consistency reliability is reported to be 0.82 for language mechanics and 0.99 for language expression. For Social Studies, this reliability is 0.87. For late sixth grade (posttest level), the internal consistency reliability for the CTBS, Form S, Level 3 Language Arts test is 0.76 for language mechanics and 0.87 for language expression. For Social Studies, this reliability is 0.89. Interlevel correlations are 0.68 (language mechanics) and 0.77 (language expression) for Language Arts and 0.73 for Social Studies.

Gates-MacGinitie Reading Tests (GMG)

The Gates-MacGinitie Reading Tests (GMG) were developed by A.I. Gates and W.H. MacGinitie and were published in 1972. They represent a series of standardized reading achievement tests for grades 4 through 6. In the present study, the Gates-MacGinitie Reading Test, Survey D, Form 1 was employed. It

It was administered at the beginning and end of the second year (pretest and posttest). This test consists of three separate scales: (1) the Speed and Accuracy Test; (2) the Vocabulary Test; and (3) the Comprehension Test. However, only two of these scales, Vocabulary and Comprehension, were utilized in this study. Further, because it was found that these scores were not sufficiently differentiated to be maintained separately, a composite total score consisting of the sum of the number right on both tests was employed.

The reported internal consistency reliability (Spearman-Brown's split-half formula) is 0.87 for the Vocabulary Test and 0.96 for the Comprehension Test. Test-retest reliability is not yet available. Evidence of validity (Gates and MacGinitie, 1972) is based on the comparison of this test with four other standardized reading tests. Coefficients of 0.78 for Vocabulary and 0.80 for Comprehension are cited for this comparison.

Results and Discussion

Stability of teachers' behavior was analyzed using the subset of 28 teachers for whom all measures were available in both years of the study. The correlations of the ratings of the teachers, between instruments and between years, were tested for significance using a two-tailed t-test, with the probability of a Type I error set at 0.01 and 26 degrees of freedom. At this level, correlations of 0.48 or higher are regarded as significant.

All of the teacher factor scores on the COR were significantly correlated across the two years. The "Kindly, Understanding" score showed a 0.77 correlation; "Systematic, Organized" was 0.78; and "Stimulating, Inventive" was 0.53. However, since half of the raters in the second year were the same as in the first, at least part of the agreement between the two years might have been due to a "carry-over" effect, in that the rating of a teacher in the second year of the study might tend to reflect the opinion formed by the same rater

in the first year. To address this issue, the COR factor scores for Year 2 were recomputed for each teacher, using only those ratings derived from raters who had not observed that particular teacher in the first year of the study. In all, three different raters were employed for this purpose, one per teacher, and scores for 23 teachers were obtained. (Five of the teachers did not have different raters in the second year). These scores were then correlated with the original first year scores, generated by a different set of observers. Again, all of the correlations were significant ($KU = 0.71$; $SO = 0.74$; $SI = 0.57$). Furthermore, they were quite similar to those obtained using all raters in both years of the study. It would appear, then that as measured by trained observers' ratings, teachers show substantial stability of behavior, even when teaching different subjects to different groups of students.

In addition, if the COR-Pupil score can be said to reflect a teacher effect, to some degree, then its significant correlation across the two different classes in the two years (0.68 for both full sets of raters; 0.65 for different raters for different years) further suggests continuity of teacher behavior, insofar as its consequences appear to be similar across different classes of students.

Ratings of teachers by the students, themselves, lead to the same conclusion. Teacher's scores on the SET, derived at the end of each year of the study, showed a significant correlation of 0.68 across the two years, although the subject area, as well as the membership of the classes, differed from year to year.

Comparing scores across instruments indicates substantial agreement between trained observers and student ratings of teachers. All correlations between the COR factor scores and SET scores within years were found to be significant at either the 0.05 or 0.01 probability level (see Table 1). In Year 1, the SET scores correlated 0.68 with COR KU, 0.54 with COR SO, and 0.47 with COR SI. In

Year 2, SET correlated 0.54 with COR KU, 0.45 with COR SO, and 0.38 with COR SI. The SET and COR Pupil scores correlated 0.51 in Year 1 and 0.40 in Year 2, another positive finding on a more indirect test of teacher (or teacher-effect) consistency.

Although the students seemed to have been randomly assigned to teachers, in proportionately distributed tri-ethnic groups, some of these correlations conceivably could have been the product of non-random assignment by ability level, in that having similar types of students with regard to achievement level might evoke in teachers similar teaching strategies and cause them to exhibit similar behavioral characteristics from year to year. However, the correlation between the achievement pretests (the CTBS in Year 1 and the GMG in Year 2) is only 0.26 (non-significant). Hence, there is no indication of initial similarity in ability level between the teachers' classes from year to year, so consistency of teacher behavior cannot be attributed to this factor.

In sum, different sets of adult and student judges, in two successive years, described these teachers with a substantial degree of consistency, even though the teachers taught different subjects in the two years. Additional data in the study, both of other kinds and from a replication sample in Daviess County, Kentucky, will cast more light on the issue. For now, it does appear that there is a considerable amount of individual stability in important aspects of teaching behavior, at least across a two-year span.

TABLE 1

Correlations Between COR Factor Scores and
SET Scores within the Same Year (N=28)

	Year 1	Year 2
	Correlations	Correlations
	SET	SET
COR KU	0.68**	0.54**
COR SO	0.54**	0.45*
COR SI	0.47*	0.38*
COR Pupil	0.51**	0.40*

*probability <.05

**probability <.01

REFERENCES

- Brophy, J. & Evertson, C. Consistency of teacher behavior across different sections of junior high math or English classes. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977.
- Brophy, J., Evertson, C., Crawford, J., King, C. and Senior, K. Stability measures of classroom process behaviors across three different contexts in second and third grade classrooms. (Research Report No. 75-1). Austin, TX.: The University of Texas Research and Development Center for Teacher Education, 1975.
- CTB/McGraw-Hill. Technical Bulletin No. 1. Comprehensive Tests of Basic Skills. Monterey, California: McGraw-Hill, 1974.
- Gates, A.I. & MacGinitie, W.H. Technical Manual, Gates-MacGinitie Reading Tests. New York: Teachers College Press, Columbia University, 1972.
- Marshall, H.H., Green, J.L., and Lawrence, M.T. Stability of teacher behaviors as measured by a broad range low-inference observational system. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1976.
- Meyers, C.L. A twelve year study of teacher success and related characteristics: Progress report, 1960-1965. Long Beach, California: California State College, 1967. (ERIC ED 037-378).
- Pfeifer, I. Comparison of verbal interaction and the goals of teachers teaching classes of different ability in eleventh grade English. Unpublished doctoral dissertation, Kent State University, 1966.
- Ryans, D.F. Characteristics of teachers. Washington, D.C.: American Council on Education, 1960.
- Shavelson, R. & Dempsey-Atwood, N. Generalizability of measures of teaching behavior. Review of Educational Research, 46, 1976, 553-611.
- Simon, A. Patterns of verbal behavior in favored and non-favored classes. Classroom Interaction Newsletter, 2 (2), 1967, 47-50.
- Veldman, D.J. Student evaluations of teaching. (RMM-10) Austin, Tx.: University of Texas Research and Development Center for Teacher Education, 1970.